

Wigham S, Rodgers J, Berney T, Le Couteur A, Ingham B, Parr JR.
[Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: a systematic review](#). *Autism: International Journal of Research and Practice* 2018. DOI:
10.1177/1362361317748245

Copyright:

Copyright © 2018 The authors. Reprinted by permission of SAGE Publications.

DOI link to article:

<https://doi.org/10.1177/1362361317748245>

Date deposited:

15/02/2018

Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: a systematic review

Sarah Wigham, Jacqui Rodgers, Tom Berney,
Ann Le Couteur, Barry Ingham & Jeremy R Parr

Word count: 5959

Declaration of Interest

None.

Correspondence

Sarah Wigham, Institute of Neuroscience, Henry Wellcome Building, Faculty of Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK.

Email: sarah.wigham@ncl.ac.uk

Funding

We are grateful to the UK autism research charity Autistica (www.autistica.org.uk) for funding this study as part of the Autism Life Course and Ageing research programme.

Abstract

Accurately diagnosing autism spectrum disorders (ASD) in adulthood can be challenging. Structured questionnaires (SQs) and diagnostic measures (DMs) are frequently used to assist case recognition and diagnosis. This study reviewed research evidence on SQs and DMs published since the National Institute for Health and Care Excellence evidence update (NICE; 2014). The Cochrane library, Medline, Embase and PsycINFO were searched. Twenty studies met inclusion criteria. Sensitivity and specificity of SQs was best for individuals with previously confirmed ASD diagnoses, and reduced in participants referred for diagnostic assessments, with discrimination of ASD from mental health conditions especially limited. For adults with intellectual disability diagnostic accuracy increased when a combination of SQs were used. Evidence suggests some utility of DMs in identifying ASD amongst clinic referrals, though specificity for diagnosis was relatively low. In mental health settings the use of a single SQ is unlikely to accurately identify adults without ASD, or differentiate ASD from mental health conditions. This is important as adults seeking an ASD diagnostic assessment are likely to have co-existing mental health conditions. Robust ASD assessment tools specifically for use in adult diagnostic health services in the presence of co-occurring mental health and neurodevelopmental disorders is a research priority.

Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: a systematic review

Abstract

Accurately diagnosing autism spectrum disorders (ASD) in adulthood can be challenging. Structured questionnaires (SQs) and diagnostic measures (DMs) are frequently used to assist case recognition and diagnosis. This study reviewed research evidence on SQs and DMs published since the National Institute for Health and Care Excellence evidence update (NICE; 2014). The Cochrane library, Medline, Embase and PsycINFO were searched. Twenty studies met inclusion criteria. Sensitivity and specificity of SQs was best for individuals with previously confirmed ASD diagnoses, and reduced in participants referred for diagnostic assessments, with discrimination of ASD from mental health conditions especially limited. For adults with intellectual disability diagnostic accuracy increased when a combination of SQs were used. Evidence suggests some utility of DMs in identifying ASD amongst clinic referrals, though specificity for diagnosis was relatively low. In mental health settings the use of a single SQ is unlikely to accurately identify adults without ASD, or differentiate ASD from mental health conditions. This is important as adults seeking an ASD diagnostic assessment are likely to have co-existing mental health conditions. Robust ASD assessment tools specifically for use in adult diagnostic health services in the presence of co-occurring mental health and neurodevelopmental disorders is a research priority.

Introduction

The importance of evaluating ASD diagnostic tools in adult populations

Research on the diagnosis of Autism Spectrum Disorder (ASD) has to date predominately focussed on childhood. Compared to the number of studies recruiting child populations, there are few studies (e.g. Lehnhardt et al., 2013; Brugha et al., 2015) specifically examining the psychometric properties of tools used in the diagnosis of suspected ASD in adulthood, for example the DISCO (Diagnostic Interview for Social and Communication Disorders; Wing et al., 2002). The tools used in the diagnosis of ASD are structured questionnaires (SQs) that are generally self-report or informant completed brief measures developed as ASD screening tools; and diagnostic measures (DMs), which are more in depth assessment tools that tend to involve semi-structured interviews and interactive tasks to inform an ASD diagnosis (National Institute for Health and Care Excellence (NICE) 2012; Scottish Intercollegiate Guidelines Network (SIGN) 2016). For adults with a suspected ASD diagnosis there are significant limitations in access to high quality diagnostic assessments. The reasons for this include both the availability of diagnostic services and psychometrically established measures (Lehnhardt et al., 2013; Powell & Acker, 2016; Brugha et al., 2015). SQs and DMs can assist clinicians making a diagnosis (NICE, 2012; SIGN, 2016) and can be used in conjunction with direct observation and other sources of information (ICD-10; APA, 2013). However childhood measures are not always suitable for adults where the presentation of ASD may be masked by maturation, learned compensatory skills and the presence of co-occurring mental health and neurodevelopmental disorders. Further, informants (particularly parents) and information about early history may be unavailable (Brugha et al., 2012; 2015).

The NICE (2014) guidelines on case recognition and diagnosis of ASD in adults

As part of the development of evidence-based clinical guidelines for the assessment and management of ASD in adults, the UK National Institute for Health and Care Excellence (NICE 2014) reviewed and published evidence on the use of SQs and DMs for case recognition and diagnosis in adults with suspected ASD. These guidelines recommended a battery of SQs and DMs that could contribute to an ASD diagnosis (see Table 1 for a summary).

In practice, clinicians and researchers use a range of SQs and DMs to aid diagnosis, not all of which were included in the NICE recommendations (Rogers et al., 2015; Rutherford et al., 2016). For example, when the NICE review (2014) was completed no diagnostic test accuracy studies were included regarding the use of the Diagnostic Interview for Social and Communication Disorders (DISCO) (Leekam et al., 2002; Wing et al., 2002) or the Developmental, Dimensional and Diagnostic Interview (3DI) (Skuse et al., 2004), in adult populations. In addition, some published SQs (usually in self-report format) have been developed to measure ASD traits that may not directly map onto diagnostic criteria, e.g. the AQ (Autism Spectrum Quotient; Baron-Cohen et al., 2001). Given the current availability of various different tools, the psychometric properties of the SQs and DMs used in clinical and research practice require further consideration (Rogers et al., 2015; Rutherford et al., 2016).

Aims of the study

The aims of this systematic review were therefore to:

- (a) Identify studies published since the NICE (2014) guidelines update on evidence for the assessment of ASD in adults that describe the psychometric properties of the SQs and DMs available to clinicians and researchers for identifying ASD in adults.
- (b) Examine the quality of the identified studies.

(c) Make recommendations about which of the measures, if any, might be most appropriate for the diagnostic assessment of ASD in adults, based on current evidence.

Table 1 around here

Method

NICE first published the adult autism clinical guideline (CG142) in 2012 (with an evidence update in 2014). This systematic literature review includes papers published between January 2013 and October 2017; and coincides with the publication of DSM-5 and the revised criteria for ASD that are used throughout this paper (APA, 2013). The review search terms and inclusion criteria are shown in Table 2.

Titles and abstracts were screened by SW with 20% screened independently by a second reviewer (TB). SW and TB reviewed full texts of selected articles independently; uncertainties about inclusion were resolved through discussion with the research team. The selection process to determine eligibility of articles for inclusion in the review is shown in Figure. 1. Data were synthesised narratively and risk of bias was assessed using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) measure for examining the quality of diagnostic studies (Whiting et al., 2011; Campbell et al., 2015). The second reviewer independently rated 20%; reviewers 1 and 2 had 96% agreement on bias ratings.

Table 2 around here

Results

3887 articles were identified after removal of duplicates (Figure 1). 83 articles were read in full and twenty articles were selected for inclusion in the review. A brief summary of the characteristics of the measures from the selected articles is provided in Table 3.

Figure 1 around here

Bias ratings are shown online in the data supplement Table (DS1) and summarized with the current review's findings for each measure. Most studies were conducted in Europe, two in Japan (Nishiyama et al., 2014; Takei et al., 2014) and four in the USA (Hus and Lord, 2014; Grodberg et al., 2014; Pugliese et al., 2015; Maddox et al., 2017). Two studies were multi-site with participants recruited from the USA, UK and Germany (Derks et al., 2017; Sappok et al., 2017). Six studies focussed on adults with intellectual disabilities (Sappok et al., 2014, 2015a, 2017; Mutsaerts et al., 2016; Derks et al 2017; Heinrich et al., 2017). Study designs were case-control, cross sectional or retrospective.

Table 3 around here

Structured questionnaires

SQs are usually self-report (but can be informant completed), do not require training for administration and have been used in epidemiological studies, and to gather information prior to a full diagnostic assessment (NICE 2014; Mutsaerts et al., 2016). Evidence for SQs available at the time of the NICE (2014) recommendations is shown in Table 1; new evidence identified in the current review from studies evaluating SQs since the NICE (2014) guideline is shown in Table 4.

Autism Spectrum Quotient (AQ) (Baron-Cohen et al., 2001)

The evidence for the psychometric properties of the AQ published since the NICE (2014) guidelines and covered in the current review is not consistent, varying according to study design and recruitment source. The sensitivity and specificity of the AQ-50 and the AQ-10 was good ($\geq 80\%$) when comparing archival clinical data from adults with ASD, against a general population group (Booth et al., 2013). However this is not equivalent to the

population of subjects presenting to ASD diagnostic assessment settings. In such settings case recognition and the need to consider both differential diagnosis and the identification of additional mental health problems can be challenging.

In a Japanese study (Nishiyama et al., 2014) recruiting individuals with a known diagnosis of ASD, and a comparison general population sample, the internal consistency and construct validity of short versions of the AQ were inconsistent. A strength of the study was that a battery of screening measures were administered to all participants, including the general population comparator group, to identify any co-occurring mental health conditions.

A further study recruiting individuals with schizophrenia (Lugnegard et al., 2015) from a psychiatric clinic, adults with ASD and individuals from the general population, found that ASD and schizophrenia were both significantly associated with a higher AQ-50 score. In addition, only the AQ version with a binary response option discriminated reliably between ASD and schizophrenia and only on one subscale (attention switching) (Lugnegard et al., 2015). There was no significant difference in AQ-50 scores between individuals with ASD and those with schizophrenia when using the 4-point response option. Participants were diagnosed prior to the study, though diagnoses of ASD and schizophrenia were confirmed in this study using the DISCO-11 (Wing et al 2002) and the Structured Clinical Interview for DSM-IV diagnosis (SCID-1; First & Gibbon 2004) respectively. However, such purposive recruitment of individuals with a defined specific diagnosis is not comparable to the use of the AQ in ASD diagnostic assessment clinical settings.

In the fourth paper (Sizoo et al., 2015), individuals attending an ASD diagnostic clinic completed abbreviated versions of the AQ prior to a clinical assessment. The comparison group again consisted of individuals from the general population. Reported specificity (over 66%) and sensitivity (57-62%) was lower than in the Booth et al (2013) study. A strength of

the study was the recruitment of participants from ASD clinics and administration of the AQ prior to their full assessment, which is similar to procedures in clinical settings.

Finally, Ashwood et al., (2016) reported findings using the subject and informant versions of the AQ-50 and the AQ-10. Participants were consecutively referred to an ASD assessment clinic, and had high rates of comorbid mental health conditions. Across both AQ versions sensitivity was above 71% but specificity was less than 38%, and the relationship between the AQ scores and the ADI-R and the ADOS-G was weak. Thus the AQ did not reliably identify those individuals who did not have ASD. The presence of mental health conditions (such as a generalised anxiety disorder) also increased the risk of false positive ASD diagnoses using the AQ10 (Ashwood et al., 2016). This study highlights the limitations of the AQ in a more ecologically valid setting compared to studies using non-clinical general population comparison groups.

Thus overall, the findings from these studies suggest that due to low levels of specificity the AQ is not a reliable indicator of which people should progress to a full ASD assessment.

Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R) (Ritvo et al., 2011)

When the RAADS-R was used in an ASD assessment service (Sizoo et al., 2015) and compared to a general population group, sensitivity was good (73%) but specificity was low (58%). This indicated some problems accurately identifying those without ASD (Sizoo et al., 2015). Limitations in this study are similar to those reporting on the AQ, including the use of a case-control design, and purposive recruitment of a general population comparison group that lacked mental health or developmental assessment. A study strength was that participants were recruited from ASD assessment services and administered the RAADS-R prior to full

ASD assessment. The low levels of specificity reported suggest the RAADS-R was not a reliable indicator of those who should progress to a full ASD assessment.

The psychometric properties of a 14-item version (the RAADS-14) were tested by Eriksson et al. (2013) in an ASD group, a clinical control group (individuals with ADHD, mood, psychotic and borderline personality disorders), and a general population group. The ASD group had significantly higher scores compared with the clinical control group. However, although both sensitivity (97%) and specificity (95%) was good with general population controls, the specificity for comparison with the clinical control group was reduced (to 46% with ADHD) (Eriksson et al., 2013). Comparison to a psychiatric control group was clinically relevant, but study limitations again included case-control design, and the lack of a consistent use of ASD and/or other mental health assessments in the control group; in addition some participants were recruited online and for these individuals diagnoses were self-reported and not confirmed.

Social Responsiveness Scale 2 (SRS-2) (Constantino & Gruber, 2005)

Full and shortened versions of the adult self-report SRS-2 were translated into Japanese (Nishiyama et al., 2014) and completed by individuals recruited from the general population and from clinical services. These versions had good internal consistency but were correlated with symptoms of psychoticism and distress as measured by the Brief Symptom Inventory (BSI) (Derogatis & Melisaratos 1983) and the K10 (Kessler et al., 2002). A strength of the study was that the general population group were administered a battery of mental health screening measures; a limitation was diagnosis of participants prior to the study.

A Japanese translation of the SRS informant version (Takei et al., 2014) demonstrated good sensitivity in both genders, and good specificity in men with ASD, compared to individuals in a control group with mental health difficulties. However, specificity in women

was reduced (61%) indicating limitations in identifying women who did not have ASD (Takei et al., 2014). A strength of this study was confirmation of diagnosis by the research team; a limitation was that only individuals in the ASD group completed a reference standard (the ADOS), the control group had no clinical assessment to exclude the presence of ASD.

Social Communication Questionnaire (SCQ) (Berument et al., 1999)

The SCQ (German version) was completed by parents or professional carers (in the majority of cases), of adults with intellectual disability and suspected ASD as part of routine clinical assessments of consecutive admissions to an intellectual disability mental health service (Sappok et al., 2015a). Interestingly, the SCQ current version (assesses the last 3 months) was found to be more effective than the lifetime version; the best combination of sensitivity (89%) and specificity (66%) was achieved by raising the SCQ cut-score for ASD from 15 to 18. When the focus was on lifetime development, specificity decreased to 48%. This finding may have been because the SCQ was completed by a professional carer who may not have known about a subject's early development, rather than a family member. 87% of included participants had a moderate/severe intellectual disability, so the findings may not generalise to adults with milder intellectual disability (Sappok et al., 2015a). A strength of the study was that no exclusions were made on grounds of neurodevelopmental or mental health comorbidities, increasing the clinical validity of the sample.

Cross-cultural validity of the SCQ current version was examined in a study recruiting adults with intellectual disability from mental health services in Germany, the UK and the USA (Sappok et al., 2017). In contrast to the previous study (Sappok et al., 2015a) a lower cut-score of 13 was found optimal and made for sensitivity of 87% and specificity of 58%. More severe levels of intellectual disability were associated with higher scores. In some cases administration of the SCQ was prior to an ASD assessment so informants completing the

measure were blind to diagnosis; however variations in administration of the SCQ in different countries may have introduced bias. SCQ scores were influenced by gender and country with males and those recruited in Germany having higher scores.

Structured questionnaires developed for use with adults with intellectual disability

Social Communication Questionnaire for Adults with Intellectual Disabilities (SCQ-AID)

(Derks et al., 2017)

International data (from Germany, the USA and UK) were collected from individuals recruited via intellectual disability mental health settings with the SCQ-AID completed by carers or researchers. The new algorithm with a cut-point of 9 demonstrated sensitivity of 81% suggesting the shortened version has clinical utility though as found by Sappok et al., (2015a, who used the original version) specificity was low (62%). A strength of the study was the international recruitment of participants; however the authors noted some differences in diagnostic process across sites. Further independent evaluation of this new measure is needed.

The Diagnostic Behavioural Assessment for ASD – Revised (DiBAS-R) (Sappok et al., 2015b)

When the DiBAS-R was completed by a relative or staff member supporting an adult accessing an intellectual disability service, rates of 81% sensitivity and specificity were found when comparing those with and without ASD (Sappok et al., 2014). These findings using this new measure are encouraging, and a strength of the study is that diagnoses were made contemporaneously therefore replicating circumstances from clinical settings. Further evaluation in other populations of adults with intellectual disability is required.

In a study of consecutive referrals to a specialist intellectual disability psychiatric clinic in Germany, DiBAS-R sensitivity was 82% and specificity 67%. Diagnostic accuracy

was better in adults with mild/moderate intellectual disability than those with severe intellectual disability (for whom specificity dropped to 34%) (Heinrich et al., 2017). A strength of the study was that the DiBAS-R was completed by carers, prior to a full diagnostic assessment, and was therefore representative of the clinical process. One study used a combination of the DiBAS-R and the Autism Checklist (ACL) and is described below (Mutsaerts et al., 2016).

Autism Checklist (ACL) (Sipes and Matson, 2014)

In a psychiatric clinic for adults with intellectual disability, Mutsaerts et al., (2016) reported that combined use of the ACL and the DiBAS-R led to improved sensitivity of 95% (when there was a positive screen on at least one measure) compared to when each measure was used alone (91%/75% respectively). Combined use also led to an increase in specificity to 88% (when there were positive screens on both measures) compared to using each measure separately (75%) (Mutsaerts et al., 2016). The inclusion of data collected from people referred to clinic was a study strength.

Table 4 around here

Diagnostic measures and observational assessments

Findings from studies of DMs and observational assessments, are shown in Table 5.

Diagnostic measures: interviews

Autism Diagnostic Interview-Revised (ADI-R) (Lord et al 1994)

The ADI-R was evaluated with adults without intellectual disability who were consecutively referred (mostly by self or a relative) to an ASD diagnosis clinic in Italy (Fusar-Poli et al., 2017). Overall the ADI-R had low sensitivity (43%) but good specificity (95%). In contrast

across the subscale domains sensitivity was acceptable (over 60%) and specificity was good (over 79%), with the exception of the restricted, repetitive and stereotyped behaviours domain (just 37%).

When the ADI-R and ADOS-G were combined sensitivity was still low (42%). The authors noted that the ADI-R may work less well with higher-functioning adults because a significant number of items rely on developmental history, and so more subtle presentations later in life may be missed (Fusar-Poli et al., 2017). It was a strength of the study that those administering the ADI-R were blind to DSM-5 based clinical consensus and that a third of participants already had a psychiatric diagnosis.

Diagnostic measures: observational

Autism Diagnostic Observation Schedule (second edition) (ADOS-2) Module 4 revised algorithm (Lord et al., 1999; Lord et al., 2012)

Hus & Lord (2014) reported sensitivity and specificity of the new revised Module 4 algorithm to be above 80% in a study including previous research participants and individuals presenting to a developmental disabilities clinic (the control group comprised individuals with a variety of DSM-IV-TR disorders). The severity scores were mostly independent of verbal IQ and race. A limitation of the study was the retrospective design, and exclusion of some individuals with a relative with ASD, so these new algorithms do require replication in other groups.

In a Dutch study of individuals with ASD and three different control groups de Bildt et al. (2016) reported a sensitivity of 61% for the ADOS Module 4. However, specificity varied according to the comparator group: 95-100% for individuals recruited from non-clinical or forensic settings, compared to 22-50% for individuals with schizophrenia (de Bildt et al., 2016). Items that were endorsed by the ASD group, but not by individuals with

schizophrenia included ‘quality of rapport’, ‘conversation’, ‘quality of social responses’ and ‘highly specific topics’. Limitations in the study design included the small sample size, and the use of archival video data with individuals being diagnosed prior to the study.

In the third study Pugliese et al., (2015) investigated individuals referred for an ASD assessment and previous research participants (many of whom met criteria for a range of diagnoses) using the ADOS-2 Module 4, and a cut-point of 8. The study found that while sensitivity was 85% (80-89% across research sites), specificity was somewhat lower at 72% (62-91%). The results were best for women and those with above or below average verbal IQ. Limitations in the study design included variability in the reference standards used to confirm a diagnosis of ASD, and omission of any description of blinding during administration of the tests.

Langmann et al., (2017) investigated the performance of the original and revised ADOS-2 Module 4 algorithms using retrospective data. Participants were adults and adolescents with suspected ASD who had been referred to specialist ASD diagnostic clinics in psychiatry services. The revised algorithm demonstrated slightly better performance overall (sensitivity 86% and specificity 80% at cut point 8) compared to the original algorithm. However specificity for adults with a diagnosis of personality disorder compared to ASD was poor (individuals with personality disorder comprised 58% of false positive cases). In addition diagnostic accuracy was less for women and older individuals. However a possible limitation to the generalisability of these findings was that less than a quarter of those recruited were female and only 25% were over age 24 years.

Fusar-Poli et al., (2017) evaluated the ADOS-2 Module 4 with adults whose IQ was over 70 and who were consecutively referred to an ASD diagnosis clinic in Italy. The majority of participants were self or relative-referrals. The original ADOS-2 Module 4 algorithm (Lord et al 2012) had sensitivity of over 86% across all domains. Sensitivity of the

revised algorithm (Hus & Lord 2014) was 87%. Specificity was 83% for the original algorithm but reduced on individual domains (63%) and was 74% on the revised algorithm. Strengths of the study included the avoidance of a case-control design, and that those administering the ADOS-2 were blind to DSM-5 based clinical consensus. In addition, and as would be expected in a clinical setting, a proportion of those recruited had existing psychiatric diagnoses (37%).

Maddox et al., (2017) found the original and revised algorithms for the ADOS-2 Module 4 (Lord et al 2012; Hus & Lord 2014) had a sensitivity of 100% while specificity was 74% and 70% respectively. A strength of the study was recruitment of participants from psychiatric settings; however study findings indicated a high number of participants with psychosis falsely met criteria for ASD and that the social-communication domain was particularly limited in discriminating between the two.

Short observational assessments

Autism Mental State Examination (AMSE) (Grodberg et al., 2012)

The AMSE was reported to have good sensitivity and specificity for discriminating ASD from other disorders using the ADOS-G or the ADI-R (Grodberg et al., 2014). Strengths of the study design included blinded administration of the measure, and the consecutive recruitment of individuals who self-referred for assessment of ASD.

Table 5 around here

Discussion

Key Findings

The two key findings of the review are that, overall, there is very limited evidence to support the use of structured questionnaires (SQs: self-report or informant completed brief measures

developed to screen for ASD) in the assessment and diagnosis of ASD in adults. Further the evidence regarding the use of diagnostic measures (DMs: more in depth assessment tools involving semi-structured interviews and interactive tasks) suggests some utility in identifying ASD amongst clinic referrals, though specificity for eventual diagnosis of ASD was still relatively low. For both types of assessment tools, more evidence is required regarding their use in diagnostic assessments. We consider that tools are useful, but only as an aid to diagnostic decision making within a broader multidisciplinary team ASD assessment. These conclusions are in keeping with the NICE (2014) guidelines that suggest using a battery of tools to support the diagnostic process.

Structured Questionnaires. The new information reported compliments but does not change the overall NICE (2014) recommendations regarding SQs. Most SQs show limited utility in identifying those who did not have ASD (Ashwood et al., 2016). This finding was evident in studies recruiting participants in clinically realistic settings (e.g. Ashwood et al., 2016) compared to studies using a case-control design with a general population comparator (e.g. Booth et al., 2013). There were particular problems differentiating ASD from schizophrenia, which may be in part a consequence of the potential overlap in symptoms (e.g. flattened affect and aspects of social-communication) and/or underlying genetic phenotype (Cross-Disorder Group of the Psychiatric Genomics, 2013). Given the low specificity, the AQ and the RAADS-R should not in our opinion be used on their own for screening and case recognition in clinical settings. Based on published data this would apply to all versions of the AQ including the abbreviated 10 item version recommended by NICE (2014). Regarding the SRS, at the time of the NICE (2014) recommendations there were no diagnostic accuracy studies that met guideline inclusion criteria. New evidence suggests the accuracy of the SRS (Japanese translation) for adults may be limited in women, and in the presence of symptoms

of psychoticism or distress as measured using screening tools (Nishiyama et al., 2014; Takei et al., 2014). Evidence for the diagnostic accuracy of the SCQ with adults was not available previously. In the current review the SCQ was found to be more effective when using a higher cut-off score in a German cohort (Sappok et al., 2015a). However when an international cohort was recruited (Germany, USA and the UK) a lower cut-score improved diagnostic accuracy (Sappok et al., 2017). Where professional carers completed the current SCQ (without reference to developmental history) for adults with intellectual disability in a clinical setting, this showed better sensitivity and specificity than the lifetime version (Sappok et al., 2015a). This highlights the potential difficulty of gathering a developmental history to support the diagnosis of ASD in adults, which has implications for the choice of diagnostic tools used. A new and shortened version (the SCQ-AID) used in an international study had good sensitivity but low specificity, however the authors note some variation in diagnostic rates and processes across countries (Derks et al., 2017). This review found no new evidence regarding the ASDI or the AAA.

Diagnostic Measures. Studies assessing the psychometric properties of DMs (such as the ADOS-2 module 4) reported some success identifying ASD among clinic referrals though limitations remained. The ADOS-G and ADI were recommended by NICE for use with adults with and without intellectual disability. New evidence shows the new ADOS-G algorithms had good sensitivity and specificity in a study with a case-control design (Hus & Lord 2014). This was to a certain extent replicated (specificity was acceptable) in individuals consecutively referred to a psychiatric clinic with a variety of diagnoses (Fusar-Poli et al., 2017). However new studies also indicated that in clinical groups, specificity was reduced for schizophrenia, personality disorder and psychosis (de Bildt et al., 2016; Maddox et al., 2017; Langmann et al., 2017). Scores were additionally found to interact with gender and IQ

although further studies are required to replicate these findings before definite conclusions can be made (Pugliese et al., 2015; Langmann et al., 2017). The ADI-R demonstrated mixed results in adult ASD clinic referrals, having good specificity but low sensitivity (Fusar-Poli et al., 2017). The AMSE showed promising evidence as a new measure; further studies in other settings and by teams independent of the authors are required (Grodberg et al., 2014).

Combinations of tools. NICE (2014) recommended using a battery of tools to support the diagnostic process. This was based on evidence from studies using SQs and DMs in isolation, and at the time the guidelines were published there were no diagnostic accuracy studies regarding use of tools in combination. In the current review only two studies evaluated the use of a combination of tools. In adults with ASD and intellectual disability, using the ACL and the DiBAS-R improved the sensitivity and specificity of these tools (Mutsaerts et al., 2016). For adults without an intellectual disability combining the ADI-R and the ADOS-G improved specificity but reduced sensitivity (Fusar-Poli et al., 2017). Sensitivity of the combined measures was lower than to using the ADOS-G alone; the authors suggest the ADI-R may be less sensitive to the more subtle presentation of ASDs diagnosed in adulthood (Fusar-Poli et al., 2017). This new evidence whereby combined use of these measures improved diagnostic effectiveness is in accordance with NICE (2014) recommendations regarding a battery of measures (Mutsaerts et al., 2016).

Strengths and limitations of the review

A strength of this review is that it includes new evidence about measures for which at the time of the NICE (2014) review publication there was no or limited evidence available (the SCQ, DiBAS-R, ACL, SRS, and the AMSE). A further strength was that only studies reporting sensitivity and specificity, or ROC analysis were included. A limitation was that the

number of studies and the sample sizes reported were relatively small, and heterogeneity in the design of studies was considerable. Additionally case-control design was highlighted as a potential source of bias for the existing evidence on tools when the NICE (2014) guidelines were published; this design limitation was also present in a number of the studies included in this review. However, the studies reported here that used more ecologically valid designs (for example, consecutive clinic referrals) also reported poor psychometric properties of most measures, particularly around specificity.

Implications

The diagnosis of ASD during adulthood is likely to pose challenges to clinicians and researchers as people's presentation may be more subtle (Brugha et al., 2015; Rogers et al., 2015). The presence of other mental health conditions, which is highly likely in individuals presenting to a specialist mental health service, challenges the accuracy of case recognition of ASD, and complicates the ability of clinicians and researchers to identify the core symptoms of ASD (Underwood et al., 2015). The risk of diagnostic overshadowing (i.e. misattributing behaviours to an existing diagnosis leading to a failure in identifying other diagnoses) as a consequence of the presence of associated psychiatric disorders may be increased (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Ford & Crewther, 2014). For many seeking a diagnostic assessment for suspected ASD in adulthood, the likely scenario will be an assessment in the context of co-occurring mental health needs, additional neurodevelopmental difficulties, social/relationship difficulties and/or problems with employment or education (NICE, 2012; Russell., et al 2016). This review highlights the limitations of the psychometric properties of existing diagnostic tools (SQs and DMs) to aid the accuracy of diagnostic practice, in adults particularly in the presence of co-occurring mental health diagnoses. The use of SQs that do not require a trained assessor may be useful

particularly in community primary and secondary health care settings (e.g. GPs and mental health services), to gather information about a possible neurodevelopmental disability (Booth et al., 2013; Allison et al., 2012). However, as already described, SQs alone should not be used to exclude further ASD assessment other than if the scores are extremely low (Brugha et al., 2012).

The studies included highlight that performance of SQs and DMs depend on the circumstances (countries, cultures and clinical setting) in which they are used and evaluated. For example, using ASD measures to identify individuals with ASD within the general population is a very different task to using the same measures in mental health and intellectual disability services, where ASD prevalence would be expected to be higher (Brugha et al., 2016). The interpretation of data derived from SQs and DMs including the values for sensitivity and specificity will be affected by both the prevalence of ASD in particular populations, and by the presence of coexisting conditions including intellectual disability and mental health conditions (Leeflang et al., 2009; Campbell et al, 2015). In addition, deciding how best to interpret the significance of particular SQ and DM scores is not without problems given the possibility that there may be cultural and/or country differences in scores (Sappok et al 2015a; 2017) in addition to the existing reported different clinical and research cut-off scores for many measures (Pugliese et al., 2015). Sappok et al., (2017) note that variations regarding the SCQ scores across countries may relate to differences in measure administration practices, ASD reference standards and referral processes. This has implications for international multi-site studies when researchers should agree cut-offs in advance (Sappok et al., 2017).

An inclusion criterion for the review was that study participants had completed a SQ or DM as part of a comprehensive ASD diagnostic assessment. However, in some studies, participants were diagnosed prior to the study and in others participants were consecutive

new referrals to a diagnostic clinic. Where participants were diagnosed prior to the study it was not always clear whether the ASD diagnosis was made in childhood or adulthood. This is important for several reasons: first, the presentation of adults diagnosed in childhood may be different to that of people referred for a first diagnosis in adulthood. Second, selecting adults with an existing diagnosis for studies of SQs and DM rather than adults referred for a suspected diagnosis may influence the sensitivity and specificity values (Leefflang et al., 2009; Campbell et al., 2015). Therefore the findings from studies in which individuals were first time referrals for an ASD assessment might arguably have more validity for real life clinical settings.

Future research

Further research is needed to investigate the accuracy, efficiency and comparative utility of all the tools identified in this review. Tools may require modification for use in adult ASD diagnosis services with individuals with or without intellectual disability whose neurodevelopmental conditions were not recognised in childhood, and for whom the risk of additional mental health comorbidities is high. Research should compare the sensitivity and specificity between tools when appropriately used in the diagnostic pathway with ecologically valid populations, such as individuals presenting to adult mental health settings. Only then can the clinical utility of individual SQs and DMs be directly compared. The strongest tools can then be developed further as needed – and others redeveloped, or discarded. New tools that have robust psychometric properties and that have been constructed specifically for use in adult diagnostic health services may be more useful than those adapted from childhood measures. In addition, clinicians and researchers should examine the utility of the combined use of measures to aid the assessment, differential diagnosis of ASD, and recognition of co-occurring neurodevelopmental and mental health conditions in the context

of an overall clinical formulation. An important next step will be investigating consensus views of adults with ASD and clinicians in different clinical contexts on the most effective, efficient and acceptable combinations of diagnostic measures. Whether SQs and DMs are sensitive to clinical change in response to interventions and/or progress over time is an area for future research (Brugha et al., 2015; Bolte & Diehl, 2013), as is the impact of gender and also culture on the psychometric properties of SQs and DMs – particularly as international collaborations are increasingly used to gather data from different settings (Pinto et al., 2014).

Regarding specific measures, no studies published since 2014 were found to have evaluated the ADOS modules 1-3 for adults with intellectual disability. No studies investigating the psychometric properties of the DISCO (Leekam et al., 2002; Wing et al., 2002) or peer-reviewed publications evaluating the 3DI (Skuse et al., 2004; Santosh et al., 2009) were identified that met the inclusion criteria for this review. Research on the clinical utility of these diagnostic measures that are used in a number of UK and European clinical settings is required.

This review did not include any existing interview measures of the more subtle presentation of the broader autism phenotype (BAP) (Parr et al., 2015). Given the possibility that presentations of ASD in adulthood may be more subtle, the potential value of interviews originally designed to identify the BAP for use as ASD diagnostic tools needs to be explored.

In conclusion, the findings of the current review compliment the NICE (2014) guidelines recommendations on the use of SQs and DMs in combination rather than in isolation. The current review suggests that SQs have limitations when used alone as screening measures and are unlikely to be reliable indicators of who should go on to have a full ASD diagnostic assessment. Future studies should examine the combined use of SQs alongside DMs for case recognition within the multidisciplinary diagnostic assessment in order to identify the most efficient and effective (clinical and cost) ways to ‘streamline’ diagnostic

practice and intervention planning in different settings for adults with suspected ASD. Research should also focus on which SQs and DMs both separately and in combination might be most appropriate for particular client groups – for example those with co-existing mental health conditions and those with intellectual disability.

References

- Allison C, Auyeung B and Baron-Cohen S. (2012) Toward brief “red flags” for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry* 51(2): 202-212.
- APA. (2013) *American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders Fifth Edition*. APA, 2013.
- Ashwood KL, Gillan N, Horder J, et al. (2016) Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological Medicine* 46(12): 2595-2604.
- Baron-Cohen S, Wheelwright S, Robinson J, et al. (2005) The adult Asperger assessment (AAA): A diagnostic method. *Journal of Autism and Developmental Disorders* 35(6): 807-819.
- Baron-Cohen S, Wheelwright S, Skinner R, et al. (2001) The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31(1): 5-17.
- Baron-Cohen S & Wheelwright S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, 34(2), 163-175.
- Barrett SL, Uljarević M, Baker EK et al. (2015). The Adult Repetitive Behaviours Questionnaire-2 (RBQ-2A): a self-report measure of restricted and repetitive behaviours. *Journal of autism and developmental disorders* 45(11):3680-92.
- Berument SK, Rutter M, Lord C, et al. (1999) Autism screening questionnaire: diagnostic validity. *The British Journal of Psychiatry* 175(5): 444-451.

552 Bolte EE and Diehl JJ. (2013) Measurement tools and target symptoms/skills used to assess
553 treatment response for individuals with autism spectrum disorder. *Journal of Autism*
554 *and Developmental Disorders* 43(11): 2491-2501.

555 Booth T, Murray AL, McKenzie K, et al. (2013) Brief report: an evaluation of the AQ-10 as a
556 brief screening instrument for ASD in adults. *Journal of Autism & Developmental*
557 *Disorders* 43(12): 2997-3000.

558 Brugha TS, Doos L, Tempier A, et al. (2015) Outcome measures in intervention trials for
559 adults with autism spectrum disorders; a systematic review of assessments of core
560 autism features and associated emotional and behavioural problems. *International*
561 *Journal of Methods in Psychiatric Research* 24(2): 99-115.

562 Brugha TS, McManus S, Smith J, et al. (2012) Validating two survey methods for identifying
563 cases of autism spectrum disorder among adults in the community. *Psychological*
564 *Medicine* 42(3): 647-656.

565 Brugha TS, Spiers N, Bankart J, et al. (2016) Epidemiology of autism in adults across age
566 groups and ability levels. *The British Journal of Psychiatry* Jul 7:bjp-p.

567 Campbell JM, Klugar M, Ding S, et al. (2015) Diagnostic test accuracy: methods for
568 systematic review and meta-analysis. *International journal of evidence-based*
569 *healthcare* 13(3): 154-162.

570 Constantino JN and Gruber CP. (2005) *Social responsiveness scale (SRS)*: Western
571 Psychological Services Los Angeles, CA.

572 Cross-Disorder Group of the Psychiatric Genomics Consortium (2013) Identification of risk
573 loci with shared effects on five major psychiatric disorders: a genome-wide analysis.
574 *The Lancet* 381(9875): 1371-1379.

575 de Bildt A, Sytema S, Meffert H, et al. (2016) The Autism Diagnostic Observation Schedule,
576 module 4: Application of the revised algorithms in an independent, well-defined,
577 Dutch sample (n = 93). *Journal of Autism and Developmental Disorders* 46(1): 21-30.

578 Derks O, Heinrich M, Brooks W, Sterkenburg P, McCarthy J, Underwood L & Sappok T
579 (2017). The Social Communication Questionnaire for adults with intellectual
580 disability: SCQ-AID. *Autism Research* 10(9): 1481–1490.

581 Derogatis LR & Melisaratos N (1983). The Brief Symptom Inventory: An introductory
582 report. *Psychological Medicine* 13(03): 595–605.

583 Eriksson JM, Andersen LMJ and Bejerot S. (2013) RAADS-14 Screen: validity of a
584 screening tool for autism spectrum disorder in an adult psychiatric population.
585 *Molecular Autism* 4(49).

586 First MB & Gibbon M . The Structured Clinical Interview for DSM-IV Axis I Disorders
587 (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-
588 II) . In: Hilsenroth MJ , Segal DL , editors . Comprehensive handbook of
589 psychological assessment, Vol 2: Personality assessment . Hoboken, NJ: Wiley &
590 Sons , 2004 .

591 Ford TC and Crewther DP. (2014) Factor Analysis Demonstrates a Common Schizoid
592 Phenotype within Autistic and Schizotypal Tendency: Implications for
593 Neuroscientific Studies. *Frontiers in Psychiatry* 5 (117).

594 Fusar-Poli L, Brondino N, Rocchetti M, Panisi C, Provenzani U, Damiani S, Politi P. (2017)
595 Diagnosing ASD in Adults Without ID: Accuracy of the ADOS-2 and the ADI-R.
596 *Journal of Autism and Developmental Disorders* 47:3370–3379.

597 Gillberg C, Gillberg C, Råstam M, et al. (2001) The Asperger Syndrome (and high-
598 functioning autism) Diagnostic Interview (ASDI): a preliminary study of a new
599 structured clinical interview. *Autism* 5(1): 57-66.

600 Grodberg D, Weinger PM, Halpern D, et al. (2014) The autism mental status exam:
601 sensitivity and specificity using DSM-5 criteria for autism spectrum disorder in
602 verbally fluent adults. *Journal of Autism & Developmental Disorders* 44(3): 609-614.

603 Grodberg D, Weinger PM, Kolevzon A, et al. (2012) Brief report: the autism mental status
604 examination: development of a brief autism-focused exam. *Journal of Autism and*
605 *Developmental Disorders* 42(3): 455-459.

606 Heinrich M, Böhm J, Sappok T. (2017) Diagnosing Autism in Adults with Intellectual
607 Disability: Validation of the DiBAS-R in an Independent Sample. *Journal of Autism*
608 *and Developmental Disorders*. <https://doi.org/10.1007/s10803-017-3336-5>.

609 Hus V and Lord C. (2014) The autism diagnostic observation schedule, module 4: revised
610 algorithm and standardized severity scores. *Journal of Autism & Developmental*
611 *Disorders* 44(8): 1996-2012.

612 ICD-10. (1992) *The ICD-10 classification of mental and behavioural disorders: clinical*
613 *descriptions and diagnostic guidelines*: Geneva: World Health Organization.

614 Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT et al. (2002). Short
615 screening scales to monitor population prevalences and trends in non-specific
616 psychological distress. *Psychological Medicine* 32(6): 959–976.

617 Langmann A, Becker J, Poustka L, Becker K, Kamp-Becker I. (2017) Diagnostic utility of
618 the autism diagnostic observation schedule in a clinical sample of adolescents and
619 adults. *Research in Autism Spectrum Disorders* 34:34-43.

- 620 Leeflang MMG, Bossuyt PMM and Irwig L. (2009) Diagnostic test accuracy may vary with
621 prevalence: implications for evidence-based diagnosis. *Journal of clinical*
622 *epidemiology* 62(1): 5-12.
- 623 Leekam SR, Libby SJ, Wing L, et al. (2002) The diagnostic interview for social and
624 communication disorders: algorithms for ICD-10 childhood autism and wing and
625 Gould autistic spectrum disorder. *Journal of Child Psychology and Psychiatry* 43(3):
626 327-342.
- 627 Lehnhardt F-G, Gawronski A, Pfeiffer K, et al. (2013) The investigation and differential
628 diagnosis of Asperger syndrome in adults. *Deutsches Arzteblatt International*
629 110(45): 755-762.
- 630 Lord C, Rutter, M, DiLavore PC et al (2012). Autism Diagnostic Observation Schedule,
631 second edition (ADOS-2) manual (part I): Modules 1–4. Torrance, CA: Western
632 Psychological Services.
- 633 Lord C, Pickles A, McLennan J, et al. (1997) Diagnosing autism: analyses of data from the
634 Autism Diagnostic Interview. *Journal of Autism and Developmental Disorders* 27(5):
635 501-517.
- 636 Lord C, Risi S, Lambrecht L, et al. (2000) The Autism Diagnostic Observation Schedule—
637 Generic: A standard measure of social and communication deficits associated with the
638 spectrum of autism. *Journal of Autism and Developmental Disorders* 30(3): 205-223.
- 639 Lord C, Rutter M, DiLavore P, et al. (1999) Autism diagnostic observation schedule (ADOS)
640 manual. *Los Angeles, CA: Western Psychological Services.*
- 641 Lord C, Rutter M and Le Couteur A. (1994) Autism Diagnostic Interview-Revised: a revised
642 version of a diagnostic interview for caregivers of individuals with possible pervasive
643 developmental disorders. *Journal of Autism and Developmental Disorders* 24(5): 659-
644 685.
- 645 Lugnegard T, Hallerback MU and Gillberg C. (2015) Asperger syndrome and schizophrenia:
646 Overlap of self-reported autistic traits using the Autism-spectrum Quotient (AQ).
647 *Nordic Journal of Psychiatry* 69(4): 268-274.
- 648 Maddox BB, Brodtkin ES, Calkins ME, Shea K, Mullan K, Hostager J, Mandell DS, Miller
649 JS. (2017). The Accuracy of the ADOS-2 in Identifying Autism among Adults with
650 Complex Psychiatric Conditions. *Journal of Autism and Developmental Disorders* 47:
651 2703-2709.
- 652 Mutsaerts CG, Heinrich M, Sterkenburg PS, et al. (2016) Screening for ASD in adults with
653 ID-moving toward a standard using the DiBAS-R and the ACL. *Journal of*
654 *Intellectual Disability Research* 60(5): 512-522.

- 655 NICE (2012) National Institute for Health and Care Excellence. Autism: Recognition,
656 Referral, Diagnosis and Management of Adults on the Autism Spectrum. (Clinical
657 Guideline 142).
- 658 NICE (2014) National Institute for Health and Care Excellence. Autism in Adults: Evidence
659 Update (59). Retrieved from www.nice.org.
- 660 NICE. (2016) National Institute for Health and Care Excellence. Surveillance report 2016 –
661 Autism spectrum disorder in adults (2012) (NICE guideline CG142). Retrieved from
662 www.nice.org
- 663 Nishiyama T, Suzuki M, Adachi K, et al. (2014) Comprehensive comparison of self-
664 administered questionnaires for measuring quantitative autistic traits in adults.
665 *Journal of Autism and Developmental Disorders* 44(5): 993-1007.
- 666 Parr JR, De Jonge MV, Wallace S, et al. (2015) New interview and observation measures of
667 the Broader Autism Phenotype: Description of strategy and reliability findings for the
668 interview measures. *Autism Research* 8(5):522-33.
- 669 Pinto D, Delaby E, Merico D, et al. (2014) Convergence of genes and cellular pathways
670 dysregulated in autism spectrum disorders. *The American Journal of Human Genetics*
671 94(5):677-94.
- 672 Powell T and Acker L. (2016) Adults' Experience of an Asperger Syndrome Diagnosis:
673 Analysis of Its Emotional Meaning and Effect on Participants' Lives. *Focus on Autism*
674 *and Other Developmental Disabilities* 31(1): 72-80.
- 675 Pugliese CE, Kenworthy L, Bal VH, et al. (2015) Replication and Comparison of the Newly
676 Proposed ADOS-2, Module 4 Algorithm in ASD Without ID: A Multi-site Study.
677 *Journal of Autism & Developmental Disorders* 45(12): 3919-3931.
- 678 Ritvo RA, Ritvo ER, Guthrie D, et al. (2011) The Ritvo Autism Asperger diagnostic scale-
679 revised (RAADS-R): A scale to assist the diagnosis of autism spectrum disorder in
680 adults: An international validation study. *Journal of Autism and Developmental*
681 *Disorders* 41(8): 1076-1089.
- 682 Rogers CL, Goddard L, Hill EL, et al. (2015) Experiences of diagnosing autism spectrum
683 disorder: a survey of professionals in the United Kingdom. *Autism* 20(7): 820-831.
- 684 Russell AJ, Murphy CM, Wilson E et al. (2016). The mental health of individuals referred for
685 assessment of autism spectrum disorder in adulthood: a clinic report. *Autism*, 20(5):
686 623-627.
- 687 Rutherford M, McKenzie K, McClure I, et al. (2016) A national study to investigate the
688 clinical use of standardised instruments in autism spectrum disorder assessment of
689 children and adults in Scotland. *Research in Autism Spectrum Disorders* 29: 93-100.

- 690 Rutter M, Le Couteur A and Lord C. (2003) Autism diagnostic interview-revised. *Los*
691 *Angeles, CA: Western Psychological Services* 29: 30.
- 692 Santosh PJ, Mandy WP, Puura K et al (2009). The construction and validation of a short form
693 of the developmental, diagnostic and dimensional interview. *European Child &*
694 *Adolescent Psychiatry*, 18(8), 521-524.
- 695 Sappok T, Diefenbacher A, Gaul I, et al. (2015a) Validity of the social communication
696 questionnaire in adults with intellectual disabilities and suspected autism spectrum
697 disorder. *American Journal on Intellectual & Developmental Disabilities* 120(3): 203-
698 214.
- 699 Sappok T, Gaul I, Bergmann T, et al. (2014) The Diagnostic Behavioral Assessment for
700 autism spectrum disorder - Revised: A screening instrument for adults with
701 intellectual disability suspected of autism spectrum disorders. *Research in Autism*
702 *Spectrum Disorders* 8(4): 362-375.
- 703 Sappok T, Gaul I, Dziobek I, et al. (2015b) Der Diagnostische Beobachtungsbogen für
704 Autismus Spektrumstörungen (DiBAS). *Psychiatrische Praxis* 42(3): 140-146.
- 705 Sappok T, Brooks W, Heinrich M, McCarthy J, Underwood L. (2017) Cross-cultural validity
706 of the social communication questionnaire for adults with intellectual developmental
707 disorder. *Journal of autism and developmental disorders* 47(2):393-404.
- 708 SIGN. (2016) Scottish Intercollegiate Guidelines Network. Assessment, diagnosis and
709 interventions for autism spectrum disorders. (SIGN 145).
- 710 Sipes M and Matson JL. (2014) Measures used to screen and diagnose ASD in young
711 children. In *Handbook of early intervention for autism spectrum disorders: Research,*
712 *policy, and practice (eds J Tarbox, DR Dixon, P Sturmey, JL Matson)*. Springer,
713 2014, 111-132.
- 714 Sizoo BB, Horwitz E, Teunisse J, et al. (2015) Predictive validity of self-report
715 questionnaires in the assessment of autism spectrum disorders in adults. *Autism* 19(7):
716 842-849.
- 717 Skuse D, Warrington R, Bishop D, et al. (2004) The developmental, dimensional and
718 diagnostic interview (3di): A novel computerized assessment for autism spectrum
719 disorders. *Journal of the American Academy of Child & Adolescent Psychiatry* 43(5):
720 548-558.
- 721 Takei R, Matsuo J, Takahashi H, et al. (2014) Verification of the utility of the social
722 responsiveness scale for adults in non-clinical and clinical adult populations in Japan.
723 *BMC Psychiatry* 14(302).

724 Underwood L, McCarthy J, Chaplin E et al (2015). Assessment and diagnosis of psychiatric
725 disorder in adults with autism spectrum disorder. *Advances in Mental Health and*
726 *Intellectual Disabilities*, 9(5): 222-229.

727 Whiting PF, Rutjes AWS, Westwood ME, et al. (2011) QUADAS-2: a revised tool for the
728 quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*
729 155(8): 529-536.

730 Wing L, Leekam SR, Libby SJ, et al. (2002) The diagnostic interview for social and
731 communication disorders: Background, inter-rater reliability and clinical use. *Journal*
732 *of Child Psychology and Psychiatry* 43(3): 307-325.

733

734

Table 1 ASD screening and diagnostic measures for adults and evidence recommended by NICE (NICE 2012; 2014)

Measure	Evidence
Structured questionnaires for case recognition	
NICE (2014) recommended the use of the following SQs (with adults without intellectual disability**) as part of a multidisciplinary diagnostic assessment, but not as diagnostic measures in isolation. The measures are easily available and training is not required	
Autism-Spectrum Quotient (AQ-10) (Allison et al, 2012)	Sensitivity/specificity (cut-off): 88%/91% (6). Conciseness and utility for case identification prior to a full diagnostic assessment in individuals when ASD was already suspected, were reported as strengths; case-control design (general population comparators) was a limitation*
Ritvo Autism Asperger Diagnostic Scale – Revised (RAADS-R) (Ritvo et al., 2011)	Sensitivity/ specificity (cut-off): 97%/ 100% (65). Evidence of reliability (test-retest, internal) and criterion validity. Case-control design (individuals with/without DSM-1V-TR conditions) was a limitation*
Adult Asperger Assessment (AAA) comprises the Autism-Spectrum Quotient (AQ) and Empathy Quotient (EQ) (Baron-Cohen 2005).	Sensitivity/ specificity (cut-off) 92% (10) Concerns about administration blinding to the reference standard (ASD assessment clinic)
Asperger Syndrome Diagnostic Interview (ASDI) (Gillberg et al., 2001)	Sensitivity/specificity (cut-off): 100%/ 91% (5/6 algorithm criteria) Some limitations in reliability and validity evidence; general population comparators
Diagnostic measures and observational assessments	
NICE (2014) recommended the use of the following DMs for adults with & without intellectual disability; though noted that the measures must be purchased and require training (Lord et al 1997; 2000; Rutter et al., 2003)	
Autism Diagnostic Interview – Revised (ADI-R) (Lord et al., 1997)	Sensitivity/specificity (cut-off): 91%/96% (communication 8; social reciprocity 10; restricted and repetitive behaviour 4)

Autism Diagnostic Observation Schedule – Generic (ADOS-G) (superseded by ADOS-2) (Lord et al., 2000)	Sensitivity/specificity (cut off): 80%/ 87% (7); 70%/ 94% (10); 90%/ 93% (13) Some evidence of reliability (inter-rater, test-retest, internal) and criterion validity
--	---

*In studies using a case-control design participants are recruited from different populations to form the ‘cases’ and ‘controls’, which does not replicate ‘real life’ clinical settings and may inflate reported diagnostic accuracy (Leeflang et al, 2009; Campbell et al, 2015).

** No case identification SQs were identified for adults with intellectual disability by NICE (2014)

Table 2 Search terms, study sources including databases, and study inclusion criteria

Searches	
Search terms	(asd or asperg\$ or autis\$) and (screen\$ or diagnos\$) and (adult\$ or adolescen\$)
Electronic database scoping searches	Health and Social Care Information Centre (HSIC) Health Management Information Centre (HMIC) Cinahl
Electronic database formal searches	Cochrane Library Medline Embase PsycINFO
Other study sources	Reference lists of included articles Google search undertaken for the names of the measures listed in NICE (2014) Lead researchers in the field contacted about recently completed studies/ forthcoming publications
Inclusion criteria	
Articles published in English	
Adult participants (where >50% of participants were over 18 years old)	
Participants had completed a structured questionnaire (SQ) or diagnostic measure (DM) as part of a comprehensive ASD diagnostic assessment	
A comprehensive ASD diagnostic assessment was considered to have taken place if the assessment was carried out by trained clinicians and incorporated (wherever possible) a developmental history e.g. from a family	

To maximise the clinical/research utility of this review, evidence of sensitivity (true positives), specificity (true negatives) or receiver operating characteristics (ROC) analyses using a diagnostic threshold was required (Campbell et al., 2015)

Exclusion criteria

Studies that only evaluated measures of one of the two core ASD symptom domains, for example, the Adult Repetitive Behaviours Questionnaire-2 (Barrett et al 2015), as a self-report measure of restricted and repetitive behaviours; and the Empathy Quotient (EQ: Baron-Cohen & Wheelwright 2004) as a measure of social cognition were excluded.

Table 3 Summary of characteristics of SQs and DMs evaluated in the studies included in this review

Structured questionnaires	
*Autism-Spectrum Quotient (AQ-10) (Allison et al, 2012)	Abridged from the 50-item version (Baron-Cohen et al., 2001). 10 items; 5 subscales: imagination, social skills, communication, attention to detail, and attention switching; cut-off = 6
*Ritvo Autism Asperger Diagnostic Scale – Revised (RAADS-R) (Ritvo et al., 2011)	80-items; 4 subscales: social relatedness, circumscribed interests, sensory–motor and language
**Social Responsiveness Scale 2 (SRS-2) (Constantino & Gruber, 2005)	Standardized 65-item four-response option measure of autism traits (with five subscales) that has both self-report and informant versions for adults.
**Social Communication Questionnaire (SCQ) (Berument et al., 1999)	The SCQ is a 40-item screening measure of current and lifetime symptoms of ASD developed from the ADI-R (Lord et al., 1994; Rutter et al., 2003). It is an informant questionnaire usually completed by parents/carers.
**Social Communication Questionnaire for Adults with Intellectual Disabilities (SCQ-AID) (Derks et al., 2017)	The SCQ-AID was developed from the SCQ (current version) by reducing the number of items to 24 and creating a new algorithm.
**Diagnostic Behavioural Assessment for ASD – Revised (DiBAS-R) (Sappok et al., 2015b)	The DiBAS-R was developed based on DSM-5 and ICD-10 ASD criteria. It is a 20-item observational screening assessment for adults with intellectual disability that can be administered by carers without training and is scored on a 4-point Likert scale.
**Autism Checklist (ACL) (Sipes and Matson, 2014)	The ACL is a 10 minute observational measure that assesses the three ICD-10 core ASD domains of social interaction, social communication, and stereotyped and restrictive behaviours. There are four items per domain scored using ordinal response options of present, partly present or not present. Training is not required; however, it is intended for administration by clinicians with ‘ASD expertise’.
Diagnostic measures and observational assessments	

*Autism Diagnostic Interview – Revised (ADI-R)	The ADI-R is a semi-structured interview conducted with parents or carers which focuses on current presentation and lifelong developmental history. The ADI-R focuses on social communication and interaction, plus restricted, repetitive and stereotyped behaviours (Lord et al 1994).
*Autism Diagnostic Observation Schedule – Generic (ADOS-G) (superseded by ADOS-2) (Lord et al., 1999; Lord et al., 2012)	The ADOS-2 is a standardized semi-structured diagnostic assessment. It is conducted through one-to-one interaction and direct observation of an individual with suspected ASD, using a range of activities and is delivered by a trained examiner. Ratings include non-verbal behaviours, mannerisms, restricted and repetitive behaviours (including repetitive speech), and aspects of social interaction (including conversational style, gestures and eye contact, imagination and creativity). Four modules are available for use with adults across a range of ability from little or no expressive language (Module 1) through to verbally fluent (Module 4). A revised set of algorithms has been published which comprise domains for social affect and restricted and repetitive behaviours. For the revised Module 4 algorithm there is a single cut-off score to differentiate ASD from non-ASD classifications in line with DSM-5, together with a measure of ASD symptom severity (Hus & Lord, 2014).

Short observational assessments

**Autism Mental State Examination (AMSE) (Grodberg et al., 2012)	The AMSE is a brief observational tool for clinicians about the presence of ASD. There is an online training package for the AMSE, which comprises eight items (including pragmatics, preoccupations, interactions, repetitive behaviours and sensitivities).
--	---

*Recommended in the NICE (2014) guidelines on case recognition and diagnosis of ASD in adults

** At the time of publication of the NICE (2014) guideline no evidence on the use of the measure with adults was available that met inclusion criteria

Table 4 Findings from the studies evaluating Structured Questionnaires

Study (country)	Structured Questionnaire	Participants (n): Recruitment source	ASD group Age years (SD) range	ASD group % Males	Findings				
					Version (cut-point)	Sensitivity	Specificity	AUC	Reliability and validity
Autism Spectrum Quotient (AQ)									
Booth et al (2013) (UK)	AQ	ASD (149): clinical & university disability service archival data. General population (134): university & social media.	15-75	NR (mostly males)	AQ 10 (6) AQ 50 (26) AQ-S (16)	80% 88% 85%	87% 80% 83%	>90%	
Nishiyama et al (2014) (Japan)	AQ	ASD (60): psychiatry & paediatric clinics. General population (3147): students & workers.	25.5 (11); 16-65	67%	AQ50 AQ28 AQ10 AQ-J21 AQ-J10 AQ-m			All versions >80% Though AQ10 & 28 sig. lower**	Alpha: AQ50, AQ-m & 28 >0.84; AQ-J10, AQ-J21, AQ20 & AQ10 = 0.64-.75. Item level discriminant validity only good for AQ28 & AQ10. Low item total correlations on all versions of AQ.

Lugnegard Hallerback & Gillberg (2015) (Sweden)	AQ	AS (51): neuro- developmental clinic. Schizophrenia (36): psychiatry outpatients. General population (49): students	27 (4)	47%				64% ASD & SP; 89% ASD & GP	AQ 50 scores significantly higher in schizophrenia and AS than general population; & higher in AS group (27) than schizophrenia (23) using binary scoring*; difference not significant on four point scale. ASD subscale scores significantly higher than GP but not significantly higher than schizophrenia except for 'attention switching'*** No gender effects except for 'imagination' (women lower).
Sizoo et al (2015) (Netherlands)	AQ	ASD (210): ASD diagnostic clinic referrals. General population (63): via social media.	39 (12.5) 18-55	76%	AQ 28 (80) AQ 10 (7)	57% 62%	70% 66%	Both 65%	ASD group alpha = AQ 28: 0.90; AQ 10: 0.72. Significantly higher scores for ASD than controls***

Ashwood et al (2016) (UK)	AQ	Suspected ASD (476) 69% co-morbidity: ASD diagnostic clinic referrals	32 (11) 18-70	75%	Self-report AQ 50 (≥26) AQ 50 (≥32) AQ 10 (≥ 6) Informant AQ 50 (≥26) AQ 50 (≥32) AQ 10 (≥ 6)	88% 71% 77% 93% 77% 79%	20% 35% 28% 14% 38% 31%	Both informant and self-report AQ 10 & 50 not significantly/ or only marginally predictive of ASD. No/weak correlation with ADOS-G/ ADI-R. GAD significantly predictive of AQ 10.	
Ritvo Autism Asperger Diagnostic Scale-Revised (RAADS-R)									
Sizoo et al (2015) (Netherlands)	RAADS-R	ASD (210): ASD diagnostic clinic referrals. General population (63): via social media.	39 (12.5) 18-55	76%	(98)	73%	58%	67%	ASD group alpha = 0.94; significantly higher scores for ASD than controls***

Eriksson et al (2013) (Sweden)	RAADS-14	ASD (77); ADHD and other psychiatric disorders (370): psychiatric outpatients or online mental health forums; general population (590): professionals via mental health lectures.	35 (11); 16-58	42%	(≥ 14)	97%	95% (GP) 64% (PC) 46% (ADHD)	99% 91% 88%	Full scale alpha = 0.9 Significantly higher total/item scores in ASD V controls*** Overlap with social anxiety (2 items). Females: significantly higher sensory reactivity scores across all groups.
Social Responsiveness Scale (SRS)									
Nishiyama et al (2014) (Japan)	SRS 2	ASD (60): psychiatry & paediatric clinics. General population (3147): students & workers.	25.5 (11); 16-65	67%	SRS2			>87%	SRS-2 (65, 30 & 11 item versions): all versions alpha >0.80. Correlations with psychological distress and psychoticism.

Takei et al (2014) (Japan)	SRS	ASD (65); psychiatric disorders (78): all psychiatry research volunteers and developmental clinic. General population (458): students & workers	27(8) 19-51	68%	65 (men) 52 (women)	84% 95%	81% (PC) 61% (PC)	90% 86%	Alpha = 0.96. Single or two factor structure. Significantly higher scores in ASD group than psychiatry group*** No correlation with IQ. Convergent validity with ADOS ($r = 0.34$).
Social Communication Questionnaire (SCQ)									
Sappok et al (2015a) (Germany)	SCQ	ID (68); ASD & ID (83: 13% mild; 87% moderate/ severe ID): psychiatry clinic for ID.	35 (11) 15-76	75%	Current (15) Current (18) Lifetime (15) Lifetime (20)	98% 89% 92% 79%	47% 66% 22% 48%	85% 70%	Current and lifetime scores higher in ASD + ID than ID**; discriminated across ID severity (except lifetime: only in mild ID).
Sappok et al (2017) (Germany, UK & USA)	SCQ	ASD & ID (220); ID (231)	All 29- 41(M)	>70%	Current (13)	87%	58%	80%	Higher scores in Germany* Higher likelihood of diagnosis in UK (OR=3)
SQs developed specifically for adults with intellectual disability									
Derks et al (2017) (Germany, UK & USA)	SCQ-AID	N=225 ASD & severe ID: 41% Germany;	Across sites 27-37 (5-12)	Across sites 70-80%	Reduced algorithm (9)	81%	62%	81%	SCQ-AID higher in ASD/ID group than ID group***

		14% USA; 31% UK In/out patient mental health services.							93% diagnostic agreement on full and reduced algorithm
Diagnostic Behavioural Assessment for ASD – Revised (DiBAS-R)									
Sappok et al (2014) (Germany)	DiBAS-R	ID (142); ASD & ID (77: 18% mild; 82% moderate- severe); ID mental health clinic	35 (12)	69%	Total (29); SCI (21); SRS (5)	88% 87% 94%	72% 65% 54%	89% 82% 87%	Two factors. Alpha: SCI = 0.91; SRS = 0.84; total = 0.91. Significantly higher total and subscale scores ASD group*** Correlations*** with SCQ, PDD-MRS & ACL.
Heinrich et al (2017) (Germany)	DiBAS-R	ID (289); ASD & ID (92; 42% mild/moderate; 58% severe); ID psychiatric clinic.	40(13)	68%		82%	67%	81%	Sensitivity/specificity & ID level: mild/moderate (79%/84%); severe (83%/34%).
Autism Checklist (ACL) & DiBAS-R									
Mutsaerts et al (2016) (Germany)	Combined ACL and DiBAS-R	84 ASD/ID: 18% mild; 82% moderate/severe . 64 ID: ID psychiatric clinic.	38 (12)	67%	ACL DiBAS-R ACL+ DiBAS-R At least one measure	91% 75% 70% 95%	75% 75% 88% 63%	85% 81%	Significantly higher scores on DiBAS-R and ACL in ASD than ID***

NR: not reported; AUC: area under curve; AS: Asperger Syndrome; GAD: generalised anxiety disorder; GP: general population; PC: psychiatric controls; SP: schizophrenia. ID: intellectual disability; SCI: Social Communication and Interaction; ADHD: attention deficit hyperactivity disorder; SRS: Stereotypy, Rigidity, and Sensory Abnormalities; AQ-J: Japanese version; AQ-m: modified response option of 1-4; PDD-MRS: Pervasive Developmental Disorder in Mental Retardation Scale; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Table 5 Findings from studies evaluating diagnostic measures and observational assessments

Study (country)	Measure	Participants (n) Recruitment source	ASD Mage: years (SD) range	ASD Male	Findings				
Diagnostic measures					Version (cut-point)	Sensitivity	Specificity	AUC	
Autism Diagnostic Interview-Revised (ADI-R)									
Fusar-Poli et al (2017) (Italy)	ADI-R	ASD (78); other psychiatric diagnoses (35). ASD diagnosis clinic consecutive referrals.	26(9) 18-55	73%	ADI-R	43%	95%	69%	Discriminant validity**
					CL (8)	69%	79%		
					SI (10)	71%	79%		
					RRB (3)	86%	37%		
					ADI-R & ADOS-G	42%	100%		
Autism Diagnostic Observation Schedule (second edition) (ADOS-2) Module 4 revised algorithm									
Hus & Lord (2014) (USA)	ADOS 2 Module 4	ASD (347); psychiatric (90): all research participants & ASD clinic referrals	21 (7) 9-55	80%	(8)	>80%	>80%		Calibrated severity score (CSS) independent of participant characteristics.

De Bildt et al (2016) (Netherlands)	ADOS Module 4	ASD (38); schizophrenia (18); psychopathy (16); non-clinical controls (21): all previous research participants	32 (11) 18-66	100%	(8) (10)	61% 53%	SP 50% SP 22% PP 100% GP >95%	66% ASD & SP. 75% ASD & SP & PP. 86% ASD & PP 88% ASD & GP	Mean total and subscale scores significantly higher for ASD than psychopathy & controls***; not compared to schizophrenia.
Pugliese et al (2015) (USA)	ADOS Module 4	ASD (253); non- ASD (68: 43 having psychiatric diagnosis): all research participants & ASD clinic referrals	19 (7) 11-61	77%		85%	72%		Significantly higher scores in ASD on all except 7 items. Majority of SA & RRB inter-item correlations significant. Sensitivity/ specificity slightly higher >16 years, VIQ <85 or >115. Cross gender sensitivity >80%; specificity lower for males (65 %) than females (86 %). CSS correlated with SRS*, negatively with SCQ*. No significant correlation with ADI or adaptive behaviour.
Langmann et al., (2017)	ADOS Module 4	ASD (165); Non ASD/other diagnosis (191)	21(7) 13-54	82%	Original (7) (10) Revised (8) (10)	82% 57% 86% 75%	83% 92% 80% 86%		Best discrimination: <18 years; lower and average IQ. Sensitivity females cut-off 10: 35% (original); 59% (revised).

Fusar-Poli et al (2017) (Italy)	ADOS 2 Module 4	ASD (78); other psychiatric diagnoses (35). ASD diagnosis clinic consecutive referrals.	26(9) 18-55	73%	Original Comm. RSI Revised	86% 90% 96% 87%	82% 63% 63% 74%	84% 81%	Discriminant validity***
					ADOS 2 & ADI-R	42%	100%		
Maddox et al (2017) (USA)	ADOS 2 Module 4	ASD (6); Psychosis (57); Mood disorder/ other (12) Outpatient psychiatry service.	31(14)	67%	Original Revised	100% 100%	74% 70%		Both ASD true positives & ASD false positives (psychosis) mean score on social communication domain = 10.5.
Observational tools to aid clinical judgement									
Autism Mental State Examination (AMSE)									
Grodberg et al (2014) (USA)	AMSE	ASD (23); psychiatric diagnosis (27): all ASD assessment clinic consecutive self-referrals	18-45	NR	DSM-5 reference ADOS reference (≥ 5)	91% 84%	93% 92%	97% 89%	Optimal cut score for those with DSM-5 or ADOS diagnosis=5. Floor effect on language and pointing items.

RRB: Restricted and Repetitive Behaviour; CL: communication & language; SI: social interaction; Comm: communication; RSI: reciprocal social interaction; SA: social affect; SP: schizophrenia; PP: psychopathy; GP: general population; VIQ: verbal IQ; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$